Fundamental gaps with approximate density functionals:
the derivative discontinuity revealed from ensemble considerations

Eli Kraisler and Leeor Kronik

Department of Materials and Interfaces, Weizmann Institute of Science, Rehovoth 76100, Israel

(Dated: September 1, 2015)

The fundamental gap is a central quantity in the electronic structure of matter. Unfortunately, the fundamental gap is not generally equal to the Kohn-Sham gap of density functional theory (DFT), even in principle. The two gaps differ precisely by the derivative discontinuity, namely, an abrupt change in slope of the exchange-correlation (xc) energy as a function of electron number, expected across an integer-electron point. Popular approximate functionals are thought to be devoid of a derivative discontinuity, strongly compromising their performance for prediction of spectroscopic properties. Here we show that, in fact, *all* exchange-correlation functionals possess a derivative discontinuity, which arises naturally from the application of ensemble considerations within DFT, without any empiricism. This derivative discontinuity can be expressed in closed form using only quantities obtained in the course of a standard DFT calculation of the neutral system. For small, finite systems, addition of this derivative discontinuity indeed results in a greatly improved prediction for the fundamental gap, even when based on the most simple approximate exchange-correlation density functional - the local density approximation (LDA). For solids, the same scheme is exact in principle, but when applied to LDA it results in a vanishing derivative discontinuity correction. This failure is shown to be directly related to the failure of LDA in predicting fundamental gaps from total energy differences in extended systems.

I. INTRODUCTION

Density-functional theory (DFT) [1–9] is the leading theoretical framework for studying the electronic properties of matter. It is based on mapping the interacting-electron system into the Kohn-Sham (KS) system of non-interacting electrons, which are subject to a common effective potential. DFT is a first principles approach, i.e. the only necessary input for the theory is the external potential, $v_{ext}(\vec{r})$, and the total number of electrons, N; no experimental data are required. In principle, the KS mapping is exact. In practice, it involves an exchange-correlation (xc) density functional, $E_{xc}[n(\vec{r})]$, whose exact form is unknown and is always approximated.

Present-day approximations within DFT already make it widely applicable to a variety of many-electron systems in physics, chemistry, and materials science [10–14]. Specifically, quantities that can be derived from the total energy of the system, notably structural and vibrational properties, can often be obtained with a satisfactory accuracy of a few percent or better. However, extraction of quantities such as the ionization potential (IP) or the fundamental gap directly from the KS eigenvalues often results in serious discrepancies with experiment (see, e.g., Refs. [15–21]).

It is by now well-known [22–25] that for the exact (but generally unknown) xc functional the highest occupied (*ho*) KS eigenvalue would equal the negative of the ionization potential of the interacting-electron system. This result is known as the ionization potential theorem. However, the KS gap, E_g^{KS}, namely the energy difference between the lowest unoccupied (*lu*) and *ho* KS eigenvalues, would still not equal the fundamental gap of the interacting system, E_g. This difference is due to a finite, spatially uniform "jump" in the KS potential, ex-

perienced as the electron number, N, crosses an integer value. This "jump" is called the derivative discontinuity, Δ, for reasons discussed in detail below. Extensive numerical investigations have shown [16, 17, 19, 26, 27] that the value of Δ associated with the exact KS potential for various systems of interest is not at all negligible in comparison to E_g^{KS}. It is commonly understood that standard semi-local xc potentials, like the local density approximation (LDA) [28–30] or generalized-gradient approximations (GGAs) [31–36], do not possess a derivative discontinuity by construction. As a result, such approximations effectively "average over" it in the vicinity of the integer point [37, 38]. Consequently, the ionization potential theorem is grossly disobeyed and in addition the fundamental gap is greatly underestimated.

For finite systems, a correct positioning of the *ho* and the *lu* KS eigenvalues is highly advantageous when describing processes like ionization, photoemission, charge transport or transfer, etc. [39–45]. But at least the ionization potential and electron affinity, and ergo the fundamental gap, which equals the difference between the two, can be calculated based on total energy differences between neutral, cation, and anion (see, e.g., [45–51])

For periodic systems, e.g., crystalline solids, this is no longer the case. Because such systems are represented by a unit cell with periodic boundary conditions, varying the number of electrons per unit cell means adding or subtracting charge from each replica of the unit cell and therefore an infinitely large charge from the system as a whole. The ensuing divergence is usually avoided by introducing a compensating uniform background to the unit cell, keeping the overall system neutral. However, this hinders the straightforward use of total energy differences for deducing fundamental gaps, as discussed, e.g., in Refs. [52, 53]. Thus, there is a clear advantage

in extracting the fundamental gap based on KS eigenvalues and other quantities arising in the calculation of the neutral system itself, without alteration of the number of electrons.

Many novel approaches have been developed within DFT for improving the accuracy of fundamental gap prediction beyond that afforded by conventional semi-local functionals, with an emphasis on applications to crystalline solids. These can be broadly divided into several categories.

Within the KS framework, significant attention has been devoted to the employment of functionals that do possess an inherent derivative discontinuity. Much of the effort involved the exact-exchange (EXX) functional, using the optimized effective potential (OEP) [6, 18, 54] approach [55–62]. More recently, novel semi-local functionals were constructed so as to mimic EXX-OEP properties [63–68].

Alternatively, it is possible to step outside the KS framework and use the generalized KS (GKS) scheme [18, 39, 69, 70], where mapping to a partially interacting electron system results in a KS-like equation that includes a non-multiplicative potential operator. This operator can reduce the magnitude of the derivative discontinuity, potentially driving it down to negligible values. Practical GKS schemes often rely on the Fock operator, or variants thereof. Such GKS calculations were performed, e.g., by employing the screened exchange approach [69, 71–75], by using global hybrid functionals [76–84, 84–86], range-separated hybrid functionals [80, 81, 87–96], and by applying a scaling correction to the Hartree and exchange functionals [97, 98]. Alternatively, one can step outside the KS scheme by introducing orbital-specific corrections, where different electrons of the KS system are subject to different potentials. This is achieved, e.g., using self-interaction correction methods [29, 99–101], DFT+U and Koopmans' compliant functionals [102–111], the LDA-1/2 method [112], Fritsche's generalized approach [113, 114], or a scissors-like operator to the KS system that affects only the vacant states [115].

A different possibility altogether is to sidestep the full charging problem by considering total energy differences arising from appropriately constructed partial charging schemes, e.g., by employing dielectric screening properties [53], by averaging the KS transition energies around the direct band-gap transition [116], or by employing perturbative curvature considerations [117].

But must conventional semi-local functionals really be abandoned as far as band gap prediction is concerned? Recently, this question has received some attention. For finite systems, Andrade and Aspuru-Guzik [118] and Gidopoulos and Lathiotakis [119] have suggested derivative-discontinuity correction schemes based on an electrostatic correction of the asymptotic potential [120]. Chai and Chen [121] derived a perturbative approach for the evaluation of the missing derivative discontinuity. In the first order, this perturbative treatment leads to the "frozen orbital approximation" result, discussed lately by

Baerends and co-workers [122]. We have shown that, contrary to conventional wisdom, in fact the KS potential derived from *any* xc functional possesses a derivative discontinuity [123], whose value emerges naturally and non-empirically from the ensemble generalization of DFT [22, 124–128]. These approaches have, so far, been applied to atoms and molecules and their potential for the solid-state remains unexplored.

Here, we derive an explicit, closed-form expression for the derivative discontinuity, Δ, of an arbitrary many-electron system studied with an arbitrary xc functional. This derivation is based on the ensemble generalization of the Hartree and xc energy terms suggested in Ref. [123]. Furthermore, Δ is expressed using only quantities associated with the neutral system, thereby avoiding alteration of the electron number. Therefore, the formalism is, in principle, applicable to both finite and periodic systems. Focusing on the latter, we explore analytically the scaling of Δ with system size. We find that while for the exact xc functional Δ must be independent of system size, for standard xc approximations like the LDA the derivative discontinuity vanishes. This failure is shown to directly related to the failure of LDA in predicting fundamental gaps from total energy differences in extended systems. These findings are demonstrated by illustrative calculations.

II. THE ENSEMBLE APPROACH

The central quantity we discuss below is the fundamental gap, E_g. It is defined as

$$E_g = I - A, \tag{1}$$

i.e., it is the difference between the ionization energy, I, and the electron affinity, A. As I and A involve removal and addition of an electron, respectively, in the following we analyse in detail the properties of a many-electron system with a varying number of electrons.

At zero temperature, the ground state of a many-electron system with a (possibly) fractional number of electrons $N = N_0 + \alpha$ ($N_0 \in \mathbb{N}$ and $\alpha \in [0,1]$) is described by an ensemble state [22]

$$\hat{\Lambda} = (1 - \alpha)|\Psi_{N_0}\rangle\langle\Psi_{N_0}| + \alpha|\Psi_{N_0+1}\rangle\langle\Psi_{N_0+1}|, \tag{2}$$

where $|\Psi_{N_0+p}\rangle$ is a pure many-electron ground state with $N_0 + p$ electrons and p is 0 or 1 [155], [156]. As a result, the ground-state energy $E(N)$ at a fractional N is a linear combination of the ground-state energies at the closest integer points:

$$E(N) = (1 - \alpha)E(N_0) + \alpha E(N_0 + 1). \tag{3}$$

Therefore, the function $E(N)$ is piecewise-linear (see Fig. 1(a) for an illustration).

The energy obtained in DFT using the KS system with the exact xc functional has to reproduce the piecewise-linear behavior. Janak's theorem [129] states that the
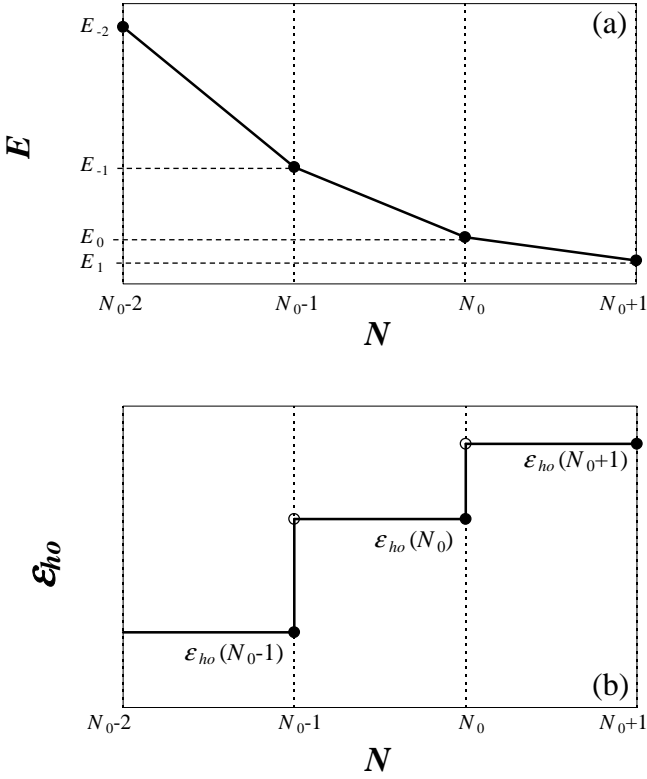
FIG. 1: A schematic plot of the total energy E (top) and the highest occupied KS energy level ε_{ho} (bottom) depending on the number of particles N for the exact xc-functional

i-th KS eigenenergy, ε_i, equals $\partial E/\partial f_i$ – the derivative of the total energy of the interacting system, E, with respect to the occupation of the i-th level, f_i. Applying this theorem, we find that with the exact xc functional the ho eigenenergy, ε_{ho}, is a stair-step function of the number of electrons, N (see Fig. 1(b)).

A non-vanishing fundamental gap in a many-electron system indicates a discontinuity in the chemical potential, namely, the cost of electron insertion and removal is different. This physical discontinuity is manifested as a mathematical discontinuity exhibited in Fig. 1. From the perspective of total energies, $I = E(N_0 - 1) - E(N_0)$ and $A = E(N_0) - E(N_0 + 1)$ are the slopes of $E(N)$ to the left and right of N_0, respectively, which are generally different from each other. The gap E_g is then the difference in these two slopes. From the perspective of KS eigenvalues, $I = -\varepsilon_{ho}(N_0)$ and $A = -\varepsilon_{ho}(N_0+1)$ (owing to the IP-theorem), and then E_g is the magnitude of the step in $\varepsilon_{ho}(N)$ at N_0 [22–25]. In other words, E_g is the discontinuity of the *derivative* of the $E(N)$ curve, or the discontinuity of the $\varepsilon_{ho}(N)$ curve itself, at N_0.

Clearly, in order to obtain the fundamental gap from total energy differences one has to calculate not only the system of interest (with $N = N_0$), but also its anion ($N = N_0 + 1$) and its cation ($N = N_0 - 1$). Figure 1(b) suggests, however, an alternative route: E_g can, at least in principle, be derived by analyzing the left and the right limits, $N \to N_0^-$ and $N \to N_0^+$, of the neutral many-electron system.

Consider now the celebrated KS equation [2], given in Hartree units by

$$\left(-\frac{1}{2}\nabla^2 + v_{KS}(\vec{r})\right)\varphi_i(\vec{r}) = \varepsilon_i\varphi_i(\vec{r}), \qquad (4)$$

where $v_{KS}(\vec{r})$ is the KS potential and $\varphi_i(\vec{r})$ are the KS orbitals. What elements of the KS equation may cause the "jump" of $\varepsilon_{ho}(N)$ at integer N, shown in Fig. 1(b)? One obvious source is simply the fact that the state named ho for $N \to N_0^+$ is *not* the same state as the one named ho for $N \to N_0^-$, due to the infinitesimal occupation of the next available energy level, whose eigenvalue is different. However, there is a more subtle second source: There is nothing to prevent the KS potential itself from exhibiting an abrupt "jump" across the integer point [18, 22, 37, 130]. Because an infinitesimal change in N across the integer point can only change the density $n(\vec{r})$ infinitesimally [37], the potential $v_{KS}(\vec{r})$ cannot "jump" by more than a spatial constant. Otherwise, in the limit of N_0 the same density would be achieved from two potentials differing by more than a constant, in direct contradiction of the Hohenberg-Kohn theorem [1]. In the complementary energy picture, the first source mentioned above – change in leading orbital – results in an abrupt slope change of the KS kinetic energy, $T_{KS} = \sum_i f_i\langle\varphi_i| - \frac{1}{2}\nabla^2|\varphi_i\rangle$. The second source – the "jump" in the KS potential – creates an abrupt slope change in the Hartree-exchange-correlation energy, E_{Hxc}.

We denote the aforementioned spatial constant by Δ and write:

$$v_{KS}^R(\vec{r}) = v_{KS}^L(\vec{r}) + \Delta. \qquad (5)$$

Here and below we use the superscripts ^{L} and ^{R} to denote quantities *immediately* to the left or to the right of the integer point N_0.

Equation (5) has two immediate consequences. Upon infinitesimal crossing of the integer point, N_0, from N_0^- to N_0^+, all the KS eigenvalues "jump" by the same quantity, i.e., $\varepsilon_i^R = \varepsilon_i^L + \Delta$. However, the KS orbitals do not exhibit any change: $\varphi_i^R(\vec{r}) = \varphi_i^L(\vec{r})$. As a special case of these statements, $\varphi_{ho}^R(\vec{r}) = \varphi_{lu}^L(\vec{r})$ and $\varepsilon_{ho}^R = \varepsilon_{lu}^L + \Delta$. These simple statements are key to the following derivation. For the gap we then obtain

$$E_g = \varepsilon_{ho}(N_0+1) - \varepsilon_{ho}(N_0) = \varepsilon_{ho}^R - \varepsilon_{ho}^L = \varepsilon_{lu}^L - \varepsilon_{ho}^L + \Delta. \qquad (6)$$

Using the definition of the KS gap, $E_g^{KS} = \varepsilon_{lu}^L - \varepsilon_{ho}^L$, we arrive at [19, 25, 37, 130]

$$E_g = E_g^{KS} + \Delta. \qquad (7)$$

The above expression is an exact result and therefore must be obeyed by results obtained from the exact KS potential. For any given approximate xc potential, however, the degree to which Eq. (6) is obeyed may vary, depending on the deviation of $\varepsilon_{ho}(N)$ from flatness, or

equivalently, on the deviation of $E(N)$ from piecewise-linearity [117, 131–139].

As already mentioned, the density is continuous across the integer point, i.e., $n^R(\vec{r}) = n^L(\vec{r})$. Because in conventional (semi-)local approximate xc functionals such as the LDA and GGAs, the xc potential is a continuous function of the density (and its gradient), it is commonly believed that there is no mathematical possibility for the KS potential to "jump" and therefore $\Delta = 0$. It is this last statement that we challenge in this work.

If the interacting-electron system has a fractional electron number, its corresponding KS system must also have a fractional electron number. Therefore, not only the ground state of the interacting system, but also the ground state of the KS system must unavoidably be described in terms of an ensemble, while still being fully described by a single KS potential. In analogy to Eq. (2), the KS ensemble state must be written in the form

$$\hat{\Lambda}^{KS} = (1-\alpha)|\Phi_{N_0}^{(\alpha)}\rangle\langle\Phi_{N_0}^{(\alpha)}| + \alpha|\Phi_{N_0+1}^{(\alpha)}\rangle\langle\Phi_{N_0+1}^{(\alpha)}|, \quad (8)$$

where $|\Phi_{N_0}^{(\alpha)}\rangle$ and $|\Phi_{N_0+1}^{(\alpha)}\rangle$ are pure KS ground states, with N_0 and $N_0 + 1$ electrons, respectively. These pure states are Slater determinants formed from the N_0 or $N_0 + 1$ occupied KS orbitals, obtained from the *same* KS potential, i.e., the two Slater determinants differ *only* in that the $N_0 + 1$ one contains one more orbital. The KS potential generating them is, generally, neither that of the pure N_0 system nor that of the pure $N_0 + 1$ system. Therefore, in contrast to the quantities $|\Psi_{N_0+p}\rangle$ used to describe the ensemble state of the interacting system, all quantities of the KS ensemble in Eq. (2) may generally change with the electron fraction, i.e., are α-dependent. We emphasize this by using the superscript (α). Ensemble-averaging the many-electron Coulomb operator $\hat{W} = \frac{1}{2}\sum_i\sum_{j\neq i}|\vec{r}_i - \vec{r}_j|^{-1}$ in the KS system, the Hartree-exchange-correlation energy functional generalizes to ensembles in the following form [123]:

$$E_{e-Hxc}[n] = (1-\alpha)E_{Hxc}[\rho_0^{(\alpha)}] + \alpha E_{Hxc}[\rho_1^{(\alpha)}]. \quad (9)$$

Here, the index $e-$ signifies that the functional is ensemble-generalized, $E_{Hxc}[\rho]$ is the pure-state Hartree-exchange-correlation energy functional, and $\rho_p^{(\alpha)}(\vec{r})$ is defined as $\rho_p^{(\alpha)}(\vec{r}) = \sum_{i=1}^{N_0+p}|\varphi_i^{(\alpha)}(\vec{r})|^2$, namely the sum of the squares of the first $N_0 + p$ KS orbitals. We stress that $\rho_p^{(\alpha)}(\vec{r})$ are auxiliary quantities that are not associated with any physical density, except when N is an integer. We further emphasize that the ensemble-generalized form of Eq. (9) is not an ansatz, but rather an inevitable consequence of employing the ensemble approach to describe a KS system of fractional number of particles. If the exact pure-state xc functional $E_{xc}[n]$ were to be inserted into Eq. (9), the ensemble-generalized total energy would have been *exactly* piecewise linear. Even then the auxiliary densities $\rho_0^{(\alpha)}$ and $\rho_1^{(\alpha)}$ need not equal the pure-state densities of N_0 - and $N_0 + 1$-systems).

The density $n(\vec{r})$ of the ensemble state is expressed in terms of $\rho_p^{(\alpha)}(\vec{r})$ as

$$n(\vec{r}) = (1-\alpha)\rho_0^{(\alpha)}(\vec{r}) + \alpha\rho_1^{(\alpha)}(\vec{r}). \quad (10)$$

To obtain $E_{e-Hxc}[n]$, we construct $\rho_0^{(\alpha)}(\vec{r})$ and $\rho_1^{(\alpha)}(\vec{r})$ from the KS orbitals as mentioned above, substitute them into the functional E_{Hxc} to obtain $E_{Hxc}[\rho_0^{(\alpha)}]$ and $E_{Hxc}[\rho_1^{(\alpha)}]$, and take the linear combination of the latter according to Eq. (9). Note that this procedure is *not* equivalent to constructing the ensemble density $n(\vec{r})$ from a linear combination of $\rho_0^{(\alpha)}(\vec{r})$ and $\rho_1^{(\alpha)}(\vec{r})$ (cf. Eq. (10)) and substituting it into E_{Hxc}, as the latter functional is not linear with respect to the density.

The generalization in Eq. (9) is applicable to *any* xc functional and makes the Hartree and the xc energy components *explicitly* dependent on the KS orbitals and on α. However, there may still remain an *implicit* non-linear dependence of $E_{e-Hxc}[n]$ on α because the KS orbitals, $\varphi_i^{(\alpha)}(\vec{r})$, may themselves change with α. Finally, note that for pure states, i.e., for $\alpha = 0$ or 1, the ensemble generalized $E_{e-Hxc}[n]$ reduces to the pure-state form $E_{Hxc}[n]$, as expected.

Because the KS potential, and specifically its behavior around an integer electron number, is central to this work, we address it here in detail. Due to the ensemble generalization of the Hartree-exchange-correlation energy functional, the KS potential is expressed as $v_{KS}(\vec{r}) = v_{ext}(\vec{r}) + v_{e-Hxc}[n](\vec{r})$, where $v_{ext}(\vec{r})$ is the external potential and $v_{e-Hxc}[n](\vec{r}) := \delta E_{e-Hxc}/\delta n$ is the ensemble-generalized Hartree-exchange-correlation potential. While deriving the latter from Eq. (9) we emphasize an unusual property of $E_{e-Hxc}[n]$: it explicitly depends on α. Therefore, the ensemble-generalized Hartree-exchange-correlation potential reads

$$v_{e-Hxc}[n](\vec{r}) = \left(\frac{\partial E_{e-Hxc}}{\partial\alpha}\right)_n\frac{\delta\alpha}{\delta n} + \left(\frac{\delta E_{e-Hxc}}{\delta n(\vec{r})}\right)_\alpha. \quad (11)$$

Since $\alpha[n] = N - \text{floor}(N)$ and $N = \int nd^3r$, we find $\delta\alpha/\delta n = 1$. Therefore, $v_{e-Hxc}[n](\vec{r})$ is a sum of two terms: $v_0[n] = (\partial E_{e-Hxc}/\partial\alpha)_n$ and $v_1[n](\vec{r}) = (\delta E_{e-Hxc}/\delta n(\vec{r}))_\alpha$. Because for fractional N the functional E_{e-Hxc} is orbital-dependent, via the quantities $\rho_p^{(\alpha)}(\vec{r})$, irrespective of the underlying xc functional, the potential $v_1[n](\vec{r})$ has to be treated with the OEP approach [6, 18, 54]. The somewhat unusual term v_0 is spatially uniform but α-dependent, and it arises from the aforementioned explicit dependence of $E_{e-Hxc}[n]$ on α.

We focus now on v_0, which can be written as

$$v_0 = \left(\frac{\partial E_{e-Hxc}}{\partial\alpha}\right)_n = \left(\frac{\partial E_{e-Hxc}}{\partial\alpha}\right)_{\{\varphi_i\}}$$
$$- \int d^3r\left(\frac{\delta E_{e-Hxc}}{\delta n(\vec{r})}\right)_\alpha\left(\frac{\partial n(\vec{r})}{\partial\alpha}\right)_{\{\varphi_i\}}. \quad (12)$$

This result is obtained by taking the partial derivative $(\partial E_{e-Hxc}/\partial\alpha)_{\{\varphi_i\}}$, followed by isolation of v_0. Using Eqs. (9) and (10) to evaluate the first and the second terms on the right-hand side of Eq. (12), respectively, we obtain

$$v_0 = E_{Hxc}[\rho_1^{(\alpha)}] - E_{Hxc}[\rho_0^{(\alpha)}] - \int d^3r |\varphi_{ho}(\vec{r})|^2 v_1[n](\vec{r}). \tag{13}$$

for $N \in (N_0, N_0 + 1]$. For $N \in (N_0 - 1, N_0]$, one has to substitute $\rho_0^{(\alpha)}$ with $\rho_{-1}^{(\alpha)}$ and $\rho_1^{(\alpha)}$ with $\rho_0^{(\alpha)}$. We stress that v_0 is a well-defined, rather than an arbitrary, potential shift. It must be taken into account for the ensemble-generalized functional, if ε_{ho} is to equal $\partial E/\partial N$, i.e., if Janak's theorem [129] is to be obeyed. The existence of a spatially uniform potential shift v_0 is in agreement with earlier studies [37, 140], which found that whereas for fractional N the KS potential is well-defined, for integer N it is defined up to a constant. The latter ambiguity in the definition of the KS potential can be removed by reaching the integer number of electrons N_0 from below (for a discussion, see [23, 25]).

Note that v_0 and $v_1(\vec{r})$ are obtained via different quantities when $N \in (N_0 - 1, N_0]$ and $N \in (N_0, N_0 + 1]$. Therefore, when approaching N_0 from the left and from the right, we generally expect to obtain different KS potentials. In other words, we expect $v_{KS}(\vec{r})$ to *change discontinuously when crossing an integer number of electrons.* As mentioned above, this discontinuity must be a spatially uniform constant, Δ (cf. Eq. (5)).

The consequences of the generalization presented above are schematically depicted in Fig. 2, based on numerical results for finite systems presented in Ref. [123]. The ensemble generalization brings $\varepsilon_{ho}(N)$ closer to the desired stair-step form: it becomes more flat for fractional N and the "jump" experienced at integer N is increased by Δ. As observed by Stein *et al.* [117], the derivative discontinuity and piecewise-linearity of the total energy are two sides of the same coin: A missing derivative discontinuity must be accompanied by deviation from piecewise linearity, and vice versa. Therefore, improvement in the description of $\varepsilon_{ho}(N)$ inevitably reflects on the total energy curve: the spurious convexity of $E(N)$ is significantly reduced, bringing it closer to the desired piecewise-linear behavior, and the abrupt change of slope near the integer points is better reproduced. Importantly, the numerical results given in Ref. [123] show that for ions of atoms and small molecules ensemble-generalization of the local spin-density approximation (LSDA) does indeed yield fundamental gaps in much better agreement with experiment than standard LSDA calculations. For example, for H_2^+, a gap of 5.80 eV predicted with LSDA is increased to 17.96 eV with ensemble-LSDA, reducing the discrepancy with respect to experiment from 70% to 8%. For C^+, LSDA predicts a gap of 0.26 eV, which is increased to 15.31 eV with ensemble-LSDA, reducing the discrepancy with experiment from 98% to 17%.

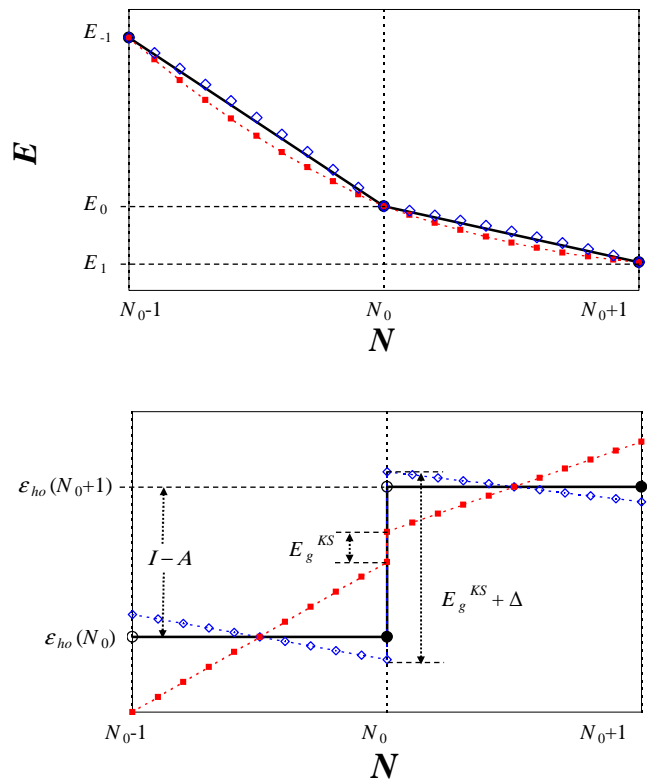Nonetheless, for an approximate xc functional the

FIG. 2: A schematic plot of the total energy E (top) and the highest occupied KS energy level ε_{ho} (bottom) as a function of the number of electrons, N, obtained with the exact functional (solid line), with an approximate functional (e.g. the LDA; squares) and with the ensemble generalization of the approximate functional (rhombi). The KS gap, E_g^{KS}, the ensemble-corrected gap, $E_g^{KS} + \Delta$ (Eq. (7)), and the exact gap, $I - A$ (Eq. (1)), are denoted on the figure.

ensemble-corrected gap $E_g = E_g^{KS} + \Delta$, still does not exactly equal $I - A$. The difference that remains is due to a deviation of $\varepsilon_{ho}(N)$ from flatness, which is attributed to the implicit non-linear dependence of an approximate $E_{e-Hxc}[n]$ on α.

III. THE DERIVATIVE DISCONTINUITY Δ

In this section, an analytical expression for the derivative discontinuity Δ is derived by taking the limits $N \to N_0^-$ and $N \to N_0^+$. First, let us introduce some notation. When $N \to N_0^+$, i.e. $\alpha \to 0^+$ (the ^{R} limit), the quantity $\rho_0^{(0)}$ is termed n_0, and the quantity $\rho_1^{(0)}$ is termed n_1. Note that because $\varphi_i^R(\vec{r}) = \varphi_i^L(\vec{r})$, these quantities are continuous when crossing N_0; for this reason they do not receive the index ^{R}. Also note that the energy $E_{e-Hxc}[n]$ and the density n were defined in Eqs. (9) and (10) for the case when $N \in [N_0, N_0 + 1]$, i.e. to the right of the point N_0. In the region $[N_0 - 1, N_0]$, which is of interest here as well, these quantities are defined similarly, substituting $\rho_0^{(\alpha)}$ with $\rho_{-1}^{(\alpha)}$ and $\rho_1^{(\alpha)}$ with $\rho_0^{(\alpha)}$. Furthermore,

when $N \to N_0^-$, i.e. $\alpha \to 1^-$ (the ^{L} limit), the quantity $\rho_{-1}^{(1)}$ is denoted n_{-1}.

We now focus on $\Delta = v_{e-Hxc}^R(\vec{r}) - v_{e-Hxc}^L(\vec{r})$, which we choose to express as $\Delta = \Delta_0 + \Delta_1$. Here

$$\Delta_0 = v_0^R - v_0^L \tag{14}$$

and

$$\Delta_1 = v_1^R(\vec{r}) - v_1^L(\vec{r}). \tag{15}$$

Reaching the point $N = N_0$ from the left, we obtain from Eq. (13)

$$v_0^L = E_{Hxc}[n_0] - E_{Hxc}[n_{-1}] - \int d^3r |\varphi_{ho}^L(\vec{r})|^2 v_1^L[n_0](\vec{r}). \tag{16}$$

Reaching N_0 from the right, we obtain similarly

$$v_0^R = E_{Hxc}[n_1] - E_{Hxc}[n_0] - \int d^3r \ |\varphi_{ho}^R(\vec{r})|^2 v_1^R[n_0](\vec{r}). \tag{17}$$

Recalling that $\varphi_{ho}^R(\vec{r}) = \varphi_{lu}^L(\vec{r})$ and using Eqs. (15) and (17), we rewrite v_0^R as

$$v_0^R = E_{Hxc}[n_1] - E_{Hxc}[n_0] - \Delta_1 - \int d^3r \ |\varphi_{lu}^L(\vec{r})|^2 v_1^L[n_0](\vec{r}). \tag{18}$$

When reaching N_0 from the left, $v_1^L[n_0](\vec{r}) = v_{Hxc}[n_0](\vec{r})$, where $v_{Hxc} = \delta E_{Hxc}/\delta n$ is the usual Hartree-exchange-correlation potential defined for the pure ground state with N_0 electrons.

Finally, subtracting Eq. (16) from Eq. (18) and using Eqs. (14) and (15), we can express the discontinuity Δ solely in terms of the ^{L} quantities, i.e. using only quantities that correspond to the system of interest, with exactly N_0 electrons. Δ is given below, suppressing now the index ^{L} for clarity:

$$\begin{aligned}\Delta = {} & E_{Hxc}[n_1] - 2E_{Hxc}[n_0] + E_{Hxc}[n_{-1}] \\ & + \int d^3r \ v_{Hxc}[n_0] \left(|\varphi_{ho}(\vec{r})|^2 - |\varphi_{lu}(\vec{r})|^2\right)\end{aligned} \tag{19}$$

Equation (19) is a key result of the current contribution. It is achieved completely from first principles, meaning that no approximations were introduced during its derivation. Because the derivation is valid for an arbitrary xc functional (exact or approximate), we conclude that *all* xc functionals possess a generally non-zero derivative discontinuity Δ, which is revealed by rigorous employment of the ensemble approach in DFT. This includes, of course, also the simplest xc approximation – the LDA, used for the computations of Ref. [123] and in calculations presented below.

Importantly, Δ as expressed in Eq. (19) is obtained using only quantities that belong to the original system of interest with N_0 electrons. Therefore its calculation does not require any alteration of the number of electrons in the system. In particular, it is also applicable to periodic systems, namely, "jellium" background charge corrections do not have to be considered in Eq. 19 because it is derived from a limit around the equilibrium point rather than from actual addition of charges.

It is well-known (see, e.g., [141]) that although the band structure of the KS system cannot be rigorously related to properties of the interacting system, it nevertheless can serve as an approximation to the charged excitation spectrum of the latter, apart from a rigid shift of the unoccupied bands with respect to the occupied ones. The corresponding shift is usually introduced empirically, or by relying on theories beyond DFT, e.g. many-body perturbation theory, and bears the name of the "scissors shift". Here, Δ provides a similar effect, with the important difference that it is derived completely within DFT.

The derivation above was performed within the OEP framework. However, it is important to note that the calculation of Δ does *not* require any actual use of the OEP formalism, but requires only simple operations of negligible numerical effort with quantities readily available from a routine DFT calculation. Actual employment of the OEP scheme is needed only for calculation of the $E(N)$ curve for fractional N.

IV. THE LIMIT OF AN INFINITELY LARGE SYSTEM

As discussed in the preceding section, Eq. (19) is applicable in principle to both finite and infinite systems. In this section, we investigate the properties of Δ for a periodic system by considering how it scales with system size as the latter approaches infinity. We obtain the analytical limiting expression and address its properties for both the exact exchange-correlation potential and the local density approximation (LDA).

Consider a many-electron system, whose external potential, $v_{ext}(\vec{r})$, is periodic in space, i.e., $v_{ext}(\vec{r} + \vec{R}) = v_{ext}(\vec{r})$, where $\vec{R}$ is a Bravais lattice vector. Neglecting surface effects, all properties of this system, including its derivative discontinuity Δ, can be obtained from the limit of a collection of M unit cells as $M \to \infty$. Let us define some terms that are useful for taking such a limit. The total number of electrons the system is MN_0, where N_0 is the (finite) number of electrons per unit cell. The electron density is $n_0(\vec{r}) = \sum_{i=1}^{MN_0} |\varphi_i(\vec{r})|^2$. The KS orbitals $\varphi_i(\vec{r})$ are, as usual, normalized to 1 when integrating over the whole system, i.e., $\int_{all} |\varphi_i(\vec{r})|^2 d^3r = 1$, where the subscript *all* denotes integration over the entire system. Therefore, $\int_{all} n_0(\vec{r}) d^3r = MN_0$ as appropriate. Integration over one unit cell yields $\int_{u.c.} n_0(\vec{r}) d^3r = N_0$ and $\int_{u.c.} |\varphi_i(\vec{r})|^2 d^3r = M^{-1} \to 0$, where the subscript *u.c.* denotes integration over one unit cell. We therefore define a renormalized KS orbital, $\bar{\varphi}_i(\vec{r}) = \sqrt{M}\varphi_i(\vec{r})$, such that $\int_{u.c.} |\bar{\varphi}_i(\vec{r})|^2 d^3r = 1$. Like the electron density, $|\bar{\varphi}_i(\vec{r})|^2$ remains finite for large M.

To assess the limiting form of Eq. (19), we first

address $E_{Hxc}^{all}[n_0 + |\varphi_{lu}|^2]$, which can be written as $E_{Hxc}^{all}[n_0 + \frac{1}{M}|\bar{\varphi}_{lu}|^2]$ using the renormalized orbitals. The Hartree-exchange-correlation energy can then be Taylor-expanded around n_0, with $1/M$ serving as the small parameter, in the form:

$$E_{Hxc}^{all}\left[n_0 + \frac{1}{M}|\bar{\varphi}_{lu}|^2\right] = E_{Hxc}^{all}[n_0] + \frac{1}{M}\int_{all}d^3r\left.\frac{\delta E_{Hxc}}{\delta n(\vec{r})}\right|_{n_0}.$$

$$|\bar{\varphi}_{lu}(\vec{r})|^2 + \frac{1}{2M^2}\int_{all}d^3r\int_{all}d^3r'\left.\frac{\delta^2 E_{Hxc}}{\delta n(\vec{r})\delta n(\vec{r}')}\right|_{n_0}.$$

$$|\bar{\varphi}_{lu}(\vec{r})|^2|\bar{\varphi}_{lu}(\vec{r}')|^2 + O\left(\frac{1}{M^3}\right) \tag{20}$$

A similar expression can be easily written for $E_{Hxc}^{all}[n_0 - |\varphi_{ho}|^2]$. Denoting the Hartree-exchange-correlation kernel by $f_{Hxc}[n](\vec{r},\vec{r}') := \delta^2 E_{Hxc}/\delta n(\vec{r})\delta n(\vec{r}')$, recognizing that $\delta E_{Hxc}/\delta n(\vec{r}) = v_{Hxc}[n](\vec{r})$, and using the renormalized orbitals $\bar{\varphi}_i(\vec{r})$ in Eq. (19), we obtain:

$$\Delta = \frac{1}{2M^2}\int_{all}d^3r\int_{all}d^3r' f_{Hxc}[n_0](\vec{r},\vec{r}')[|\bar{\varphi}_{lu}(\vec{r})|^2|\bar{\varphi}_{lu}(\vec{r}')|^2$$

$$+ |\bar{\varphi}_{ho}(\vec{r})|^2|\bar{\varphi}_{ho}(\vec{r}')|^2] + O\left(\frac{1}{M^3}\right) \tag{21}$$

The Hartree-exchange-correlation kernel $f_{Hxc}[n](\vec{r},\vec{r}')$ can be written as a sum of the Hartree and xc components: $f_{Hxc}[n](\vec{r},\vec{r}') = f_H[n](\vec{r},\vec{r}') + f_{xc}[n](\vec{r},\vec{r}')$, where $f_H[n_0](\vec{r},\vec{r}') = \delta^2 E_H/\delta n(\vec{r})\delta n(\vec{r}') = 1/|\vec{r}-\vec{r}'|$. Then, the Hartree-related term of the derivative discontinuity can be expressed as

$$\Delta_H = \frac{1}{2M^2}\sum_j\int_{all}d^3r|\bar{\varphi}_j(\vec{r})|^2 V_{Hj}(\vec{r}), \tag{22}$$

where j stands for ho or lu and $V_{Hj}(\vec{r}) = \int_{all}d^3r'|\bar{\varphi}_j(\vec{r}')|^2|\vec{r}-\vec{r}'|^{-1}$. In the limit of large M (and neglecting the diverging term because the "jellium" background is irrelevant, as explained above), both $|\bar{\varphi}_j(\vec{r})|^2$ and $V_{Hj}(\vec{r})$ are periodic and remain finite as $M \to \infty$. Therefore the integration can be performed over a unit cell, yielding:

$$\Delta_H = \frac{1}{2M}\sum_j\int_{u.c.}d^3r|\bar{\varphi}_j(\vec{r})|^2 V_{Hj}(\vec{r}). \tag{23}$$

Therefore, the Hartree-related terms decay as M^{-1} and vanish for the periodic solid.

The scaling of the xc contribution, Δ_{xc}, is obviously much more interesting and it is here that the particular choice of the xc functional is crucial. For the exact xc functional, Δ_{xc} is generally expected to be non-vanishing, because $f_{xc}[n_0](\vec{r},\vec{r}')$ is known to exhibit divergence (see, e.g., the discussion in Ref. [142, 143], and references therein). The nature of the singularity in the exact xc functional, therefore, must be such that for a periodic solid Δ_{xc} obtained from Eq. (21) is the exact
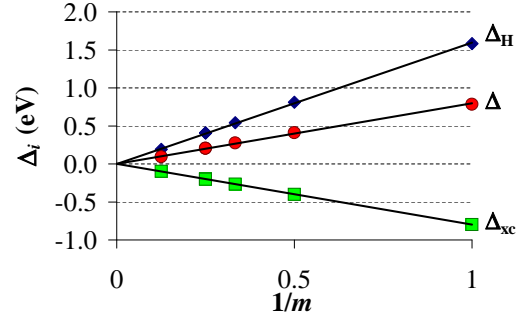

FIG. 3: The derivative discontinuity Δ, as well as its Hartree- and xc-components (Δ_H and Δ_{xc}, respectively), as a function of $1/m$, where m is the number of primitive unit cells in the supercell, for GaAs

one. Namely, the scaling for Δ_{xc} with M as $M \to \infty$ should be $\sim M^0$. In parallel, the xc potential v_{xc} must scale as $\sim M^0$, and the xc energy E_{xc} as $\sim M^1$.

Unfortunately, this is not the case for simple functionals such as the LDA. In the LDA, the xc kernel can be expressed as $f_{xc}(\vec{r},\vec{r}') = g_{xc}(\vec{r})\delta(\vec{r}-\vec{r}')$, where $g_{xc}(\vec{r})$ is a function of the density (and therefore periodic in a periodic system). As a result, the xc-related terms in Eq. (21) simplify to $\frac{1}{2M}\int_{u.c.}d^3r|\bar{\varphi}_j(\vec{r})|^4 g_{xc}(\vec{r})$, i.e. they too decay as M^{-1}. Therefore, in the LDA approximation, for infinite systems $\Delta \sim M^{-1} \to 0$.

As a practical illustration of the above analysis, we used Eq. (19) to evaluate Δ_H and Δ_{xc} in practice, focusing on GaAs as a prototypical example. Briefly, all electronic structure calculations were performed using the real-space PARSEC package [144–147], while employing periodic boundary conditions [148, 149]. We used the Perdew-Zunger parameterization [29] of LDA with norm-conserving norm-conserving [150] Troullier-Martins pseudopotentials [157].

To investigate the dependence of Δ_H, Δ_{xc}, and Δ on the system size, we calculated these three quantities for increasingly large GaAs supercells of 1x1x1, 1x1x2, 1x1x3, 1x1x4, 1x2x2, and 2x2x2 primitive unit cells. Eq. (19) was then used under the assumption that φ_{ho} and φ_{lu} can be taken as the highest occupied and lowest unoccupied orbitals, respectively, of the supercell. [158] In other words, the supercell is treated as a *finite but topologically periodic* system and therefore approaches the bulk limit as the number of primitive unit cells, m, approaches infinity.

The results are given in detail in Fig. 3. Clearly Δ_H, Δ_{xc}, and their sum, Δ, are indeed all linear with $1/m$ and vanish in the limit of a large enough supercell. As the LDA-KS gap remains a constant ~ 0.6 eV for all m, in the bulk limit the corrected LDA gap simply approaches the uncorrected one. Similar trends were obtained for several other prototypical semiconductors (e.g., Si, Ge, InP) and are not shown for brevity.

What is the physical origin for the apparent failure of the ensemble-correction for LDA (and indeed for any

semi-local functional) in the limit of a periodic solid? To understand it, consider again the results of Sec. II, in particular Fig. 2. As shown there, the ensemble correction strongly reduces the curvature of the total energy versus particle number curve. This greatly assists in bringing the fundamental gap deduced from eigenvalue differences closer to the one deduced from total energy differences. That this would also help improve agreement with experiment hinges on the assumption that the fundamental gap deduced from total energy differences is close to the experimental one. As discussed in Sec. II, for small and mid-size objects extensive numerical experience shows that this is often the case (see, e.g., [45–51]). But for the infinite limit, it is in fact known that with the LDA, whose xc kernel is not singular, the fundamental gap deduced from total energy differences corresponds poorly to experiment and simply approaches the KS gap [151] (For a similar reason, gaps deduced from time-dependent LDA also reduce to the LDA ones in the solid state limit – see [142, 152], and references therein).

Therefore, the correction corresponding to LDA should indeed vanish. From that perspective one could argue that in the solid-state limit the ensemble correction scheme "fails successfully" for LDA, as it yields precisely what it was built to deliver – consistency between total energy differences and eigenvalue differences (both of which are, alas, equally wrong). Complementarily, several studies have shown that in the bulk limit the LDA total energy versus particle number curve is piecewise-linear even without ensemble corrections, albeit with the wrong slope [136, 153]. Also from this perspective, Δ must vanish, as there is no curvature to reduce.

From yet a different perspective, mathematically the difficulty arises because the ho and lu orbitals are extremely delocalized, whereas the LDA xc kernel is extremely localized. This advocates the importance of ultra-non-local kernels. But in lieu of developing new functionals, another possibility is to localize the ho and lu orbitals. One such localization procedure is the above-mentioned dielectric screening based one suggested by Chan and Ceder [53] and others may be envisaged. In fact, one could argue that use of a small supercell as in Fig. 3 is, loosely speaking, a form of (uncontrolled) localization. Indeed if one were to take the results for the single unit cell literally, one would obtain $\Delta=0.78$ eV which would suggest a satisfying (but deceptive) agreement between the fundamental gap, $E_g^{KS} + \Delta = 1.39$ eV and the experimental fundamental gap value, 1.51 eV [154]. A similar behavior is obtained for other semiconductors as well. This suggests that controlled, physically justified localization procedures may prove to be key to systematic gap predictions even within LDA.

V.   CONCLUSIONS

In this article, we have revisited the issue of the derivative discontinuity from an ensemble-DFT perspective.

We have shown much of the deviation of approximate functionals from piecewise linearity is in fact due to the lack of an ensemble treatment. We have used this to show that *all* exchange-correlation functionals possess a derivative discontinuity, which arises naturally from the application of ensemble considerations within DFT, without any empiricism or any further approximations beyond the choice of the xc functional. We then expressed this derivative discontinuity in closed form using only quantities obtained in the course of a standard DFT calculation of the neutral system. We showed that for small, finite systems, addition of this derivative discontinuity indeed results in a greatly improved prediction for the fundamental gap, even when based on the most simple approximate exchange-correlation density functional - the local density approximation (LDA). We then discussed the limit of an infinitely large system, so as to approach the solid-state limit. We found that the same scheme is exact in principle, but results in a vanishing derivative discontinuity correction when applied to semi-local functionals. This failure was shown to be directly related to the failure of semi-local functionals in predicting fundamental gaps from total energy differences in extended systems. Last, we discussed possible future remedies, especially usage of localization schemes.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[3] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1989).

[4] R.M. Dreizler and E.K.U. Gross, *Density Functional Theory* (Springer Verlag, Berlin, 1990).

[5] C. Fiolhais, F. Nogueira, and M. A. Marques, eds., *A Primer in Density Functional Theory* (Springer, 2003), vol. 620 of *Lectures in Physics*.

[6] E. Engel and R. Dreizler, *Density Functional Theory: An Advanced Course* (Springer, 2011).

[7] K. Burke, J. Chem. Phys. **136**, 150901 (2012).

[8] K. Capelle, Braz. J. Phys. **36**, 1318 (2006).

[9] N. Argaman and G. Makov, Am. J. Phys. **68**, 69 (2000).

[10] R. Martin, *Electronic Structure* (Cambridge Unviersity Press, 2004).

[11] J. Hafner, Acta Mater. **48**, 71 (2000).

[12] E. Kaxiras, *Atomic and Electronic Structure of Solids* (Cambridge University Press, 2003).

[13] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (Wiley, 2005).

[14] D. Sholl and J. Steckel, *Density Functional Theory: A Practical Introduction* (Wiley, 2011).

[15] S. G. Louie, in *Topics in Computational Materials Science*, edited by C. Fong (World Scientific, Singapore, 1997), pp. 96–142.

[16] G. K.-L. Chan, J. Chem. Phys. **110**, 4710 (1999).

[17] M. Allen and D. Tozer, Mol.Phys. **100**, 433 (2002).

[18] S. Kümmel and L. Kronik, Rev. Mod. Phys. **80**, 3 (2008).

[19] A. M. Teale, F. de Proft, and D. J. Tozer, J. Chem. Phys. **129**, 044110 (2008).

[20] S. Refaely-Abramson, R. Baer, and L. Kronik, Phys. Rev. B **84**, 22 (2011).

[21] X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).

[22] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).

[23] M. Levy, J. P. Perdew, and V. Sahni, Phys. Rev. A **30**, 2745 (1984).

[24] C. Almbladh and U. Von Barth, Phys. Rev. B **31**, 3231 (1985).

[25] J. P. Perdew and M. Levy, Phys. Rev. B **56**, 16021 (1997).

[26] R. Godby, M. Schlüter, and L. Sham, Phys. Rev. B **36**, 6497 (1987).

[27] R. Godby, M. Schlüter, and L. Sham, Phys. Rev. B **37**, 10159 (1988).

[28] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).

[29] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[30] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[31] J. P. Perdew and Y. Wang, Phys. Rev. B **33**, 8800 (1986).

[32] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[33] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[34] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[35] Z. Wu and R. Cohen, Phys. Rev. B **73**, 235116 (2006).

[36] P. Haas, F. Tran, P. Blaha, and K. Schwarz, Phys. Rev. B **83**, 205117 (2011).

[37] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[38] D. J. Tozer and N. C. Handy, J. Chem. Phys. **109**, 10180 (1998).

[39] L. Kronik, T. Stein, S. Refaely-Abramson, and R. Baer, J. Chem. Theory Comp. **8**, 1515 (2012).

[40] L. Kronik and S. Kümmel, *Gas-Phase Valence-Electron Photoemission Spectroscopy Using Density Functional Theory* in Topics of Current Chemistry: First Principles Approaches to Spectroscopic Properties of Complex Materials, to be published.

[41] S. Y. Quek, H. J. Choi, S. G. Louie, and J. B. Neaton, Nano Lett. **9**, 3949 (2009).

[42] A. Dreuw and M. Head-Gordon, J. Am. Chem. Soc. **126**, 4007 (2004).

[43] D. J. Tozer, J. Chem. Phys. **119**, 12697 (2003).

[44] T. Stein, L. Kronik, and R. Baer, J. Am. Chem. Soc. **131**, 2818 (2009).

[45] B. Kaduk, T. Kowalczyk, and T. Van Voorhis, Chem. Rev. **112**, 321 (2012).

[46] S. Öğüt, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. Lett. **79**, 1770 (1997).

[47] L. Kronik, R. Fromherz, E. Ko, G. Ganteför, and J. R. Chelikowsky, Nature materials **1**, 49 (2002).

[48] M. Moseler, B. Huber, H. Häkkinen, U. Landman, G. Wrigge, M. Astruc Hoffmann, and B. v. Issendorff, Phys. Rev. B **68**, 165413 (2003).

[49] H.-C. Weissker, J. Furthmüller, and F. Bechstedt, Phys. Rev. B **69**, 115310 (2004).

[50] E. Kraisler, G. Makov, and I. Kelson, Phys. Rev. A **82**, 042516 (2010).

[51] U. Argaman, G. Makov, and E. Kraisler, Phys. Rev. A **88**, 042504 (2013).

[52] S. Sharma, J. Dewhurst, N. Lathiotakis, and E. K. U. Gross, Phys. Rev. B **78**, 201103 (2008).

[53] M. Chan and G. Ceder, Phys. Rev. Lett. **105**, 196403 (2010).

[54] T. Grabo, T. Kreibich, and E. K. U. Gross, Mol. Eng. **7**, 27 (1997).

[55] D. Bylander and L. Kleinman, Phys. Rev. B **54**, 7891 (1996).

[56] M. Städele, J. Majewski, P. Vogl, and A. Görling, Phys. Rev. Lett. **79**, 2089 (1997).

[57] M. Städele, M. Moukara, J. Majewski, P. Vogl, and A. Görling, Phys. Rev. B **59**, 10031 (1999).

[58] R. Magyar, A. Fleszar, and E. K. U. Gross, Phys. Rev. B **69**, 045111 (2004).

[59] M. Grüning, A. Marini, and A. Rubio, J. Chem. Phys. **124**, 154108 (2006).

[60] M. Grüning, A. Marini, and A. Rubio, Phys. Rev. B **74**, 161103 (2006).

[61] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).

[62] P. Rinke, M. Winkelnkemper, A. Qteish, D. Bimberg, J. Neugebauer, and M. Scheffler, Phys. Rev. B **77**, 075202 (2008).

[63] A. D. Becke and E. R. Johnson, J. Chem. Phys. **124**, 221101 (2006).

[64] F. Tran, P. Blaha, and K. Schwarz, J. Phys.: Condens. Matter **19**, 196208 (2007).

[65] F. Tran and P. Blaha, Phys. Rev. Lett. **102**, 226401 (2009).

[66] M. Kuisma, J. Ojanen, J. Enkovaara, and T. T. Rantala, Phys. Rev. B **82**, 115106 (2010).

[67] R. Armiento and S. Kümmel, Phys. Rev. Lett. **111**, 036402 (2013).

[68] A. Laref, A. Altujar, and S. Luo, Eur. Phys. J. B **86**, 475 (2013).

[69] A. Seidl, A. Görling, P. Vogl, J. Majewski, and M. Levy, Phys. Rev. B **53**, 3764 (1996).

[70] R. Baer, E. Livshits, and U. Salzner, Annu. Rev. Phys. Chem. **61**, 85 (2010).

[71] S. Picozzi and A. Continenza, Phys. Rev. B **61**, 4677 (2000).

[72] C. B. Geller, W. Wolf, S. Picozzi, A. Continenza, and R. Asahi, Appl. Phys. Lett. **79**, 368 (2001).

[73] C. Stampfl, W. Mannstadt, R. Asahi, and A. J. Freeman, Phys. Rev. B **63**, 155106 (2001).

[74] J. Robertson, K. Xiong, and S. J. Clark, Phys. Status Solidi (B) **243**, 2054 (2006).

[75] B. Lee and L.-W. Wang, Phys. Rev. B **73**, 153309 (2006).

[76] R. Dovesi, R. Orlando, C. Roetti, C. Pisani, and V. Saunders, Phys. Status Solidi (B) **217**, 63 (2000).

[77] T. Bredow and A. Gerson, Phys. Rev. B **61**, 5194 (2000).

[78] J. Muscat, A. Wander, and N. M. Harrison, Chem. Phys. Lett. **342**, 397 (2001).

[79] F. Corà, M. Alfredsson, G. Mallia, D. S. Middlemiss, W. C. Mackrodt, R. Dovesi, and R. Orlando, Struct. Bonding **113**, 171 (2004).

[80] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Angyán, J. Chem. Phys. **124**, 154709 (2006).

[81] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Angyán, J. Chem. Phys. **125**, 249901 (2006).

[82] P. G. Moses and C. G. Van de Walle, Appl. Phys. Lett. **96**, 021908 (2010).

[83] P. G. Moses, M. Miao, Q. Yan, and C. G. Van de Walle, J. Chem. Phys. **134**, 084703 (2011).

[84] N. Sai, P. F. Barbara, and K. Leung, Phys. Rev. Lett. **106**, 226403 (2011).

[85] M. Jain, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. Lett. **107**, 216806 (2011).

[86] J. E. Moussa, P. A. Schultz, and J. R. Chelikowsky, J. Chem. Phys. **136**, 204117 (2012).

[87] J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, J. Chem. Phys. **123**, 174101 (2005).

[88] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, J. Chem. Phys. **125**, 224106 (2006).

[89] I. C. Gerber, J. G. Angyán, M. Marsman, and G. Kresse, J. Chem. Phys. **127**, 054101 (2007).

[90] H. R. Eisenberg and R. Baer, Phys. Chem. Chem. Phys. **11**, 4674 (2009).

[91] S. J. Clark and J. Robertson, Phys. Status Solidi (B) **248**, 537 (2011).

[92] T. M. Henderson, J. Paier, and G. E. Scuseria, Phys. Status Solidi (B) **248**, 767 (2011).

[93] L. Schimka, J. Harl, and G. Kresse, J. Chem. Phys. **134**, 024116 (2011).

[94] M. J. Lucero, T. M. Henderson, and G. E. Scuseria, J. Phys.: Condens. Matter **24**, 145504 (2012).

[95] T. Stein, H. Eisenberg, L. Kronik, and R. Baer, Physical Review Letters **105**, 266802 (2010).

[96] S. Refaely-Abramson, S. Sharifzadeh, M. Jain, R. Baer, J. B. Neaton, and L. Kronik, Phys. Rev. B **88**, 081204 (2013).

[97] X. Zheng, A. J. Cohen, P. Mori-Sánchez, X. Hu, and W. Yang, Phys. Rev. Lett. **107**, 026403 (2011).

[98] X. Zheng, T. Zhou, and W. Yang, J. Chem. Phys. **138**, 174105 (2013).

[99] A. Svane and O. Gunnarsson, Phys. Rev. Lett. **65**, 1148 (1990).

[100] R. Heaton, J. Harrison, and C. Lin, Phys. Rev. B **28**, 5992 (1983).

[101] A. Filippetti and N. Spaldin, Phys. Rev. B **67**, 125109 (2003).

[102] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys.: Condens. Matter **9**, 767 (1997).

[103] M. Cococcioni and S. D. Gironcoli, Phys. Rev. B **71**, 035105 (2005).

[104] A. Janotti, D. Segev, and C. Van de Walle, Phys. Rev. B **74**, 045202 (2006).

[105] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008).

[106] S. Lany and A. Zunger, Phys. Rev. B **80**, 085202 (2009).

[107] M. Forti, P. Alonso, P. Gargano, and G. Rubiolo, Procedia Materials Science **1**, 230 (2012).

[108] A. N. Andriotis, G. Mpourmpakis, S. Lisenkov, R. M. Sheetz, and M. Menon, Phys. Status Solidi (B) **250**, 356 (2013).

[109] I. Dabo, A. Ferretti, N. Poilvert, Y. Li, N. Marzari, and M. Cococcioni, Phys. Rev. B **82**, 115121 (2010).

[110] I. Dabo, A. Ferretti, C.-H. Park, N. Poilvert, Y. Li, M. Cococcioni, and N. Marzari, Phys. Chem. Chem. Phys. **15**, 685 (2013).

[111] I. Dabo, A. Ferretti, G. Borghi, N. Nguyen, N. Poilvert, C.-H. Park, M. Cococcioni, and N. Marzari, Psi-k Newsletter **119**, 1 (2013).

[112] L. Ferreira, M. Marques, and L. K. Teles, Phys. Rev. B **78**, 125116 (2008).

[113] L. Fritsche, Physica B **172**, 7 (1991).

[114] I. N. Remediakis and E. Kaxiras, Phys. Rev. B **59**, 5536 (1999).

[115] H. Mera and K. Stokbro, Phys. Rev. B **79**, 125109 (2009).

[116] P. Scharoch and M. Winiarski, Comput. Phys. Commun. **184**, 2680 (2013).

[117] T. Stein, J. Autschbach, N. Govind, L. Kronik, and R. Baer, J. Phys. Chem. Lett. **3**, 3740 (2012).

[118] X. Andrade and A. Aspuru-Guzik, Physical Review Letters **107**, 183002 (2011).

[119] N. I. Gidopoulos and N. N. Lathiotakis, J. Chem. Phys. **136**, 224109 (2012).

[120] R. V. Leeuwen and E.-J. Baerends, Phys. Rev. A **49**, 2421 (1994).

[121] J.-D. Chai and P.-T. Chen, Phys. Rev. Lett. **110**, 033002 (2013).

[122] E. Baerends, O. Gritsenko, and R. van Meer, Phys. Chem. Chem. Phys. **15**, 16408 (2013).

[123] E. Kraisler and L. Kronik, Phys. Rev. Lett. **110**, 126403 (2013).

[124] E. H. Lieb, Int. J. Quantum Chem. **24**, 243 (1983).

[125] R. van Leeuwen, Adv. Quantum Chem. **43**, 24 (2003).

[126] D. P. Joubert, Int. J. Quantum Chem. **113**, 1076 (2013).

[127] M. Hellgren and E. K. U. Gross, J. Chem. Phys. **136**, 114102 (2012).

[128] M. Hellgren and E. K. U. Gross, Phys. Rev. A **88**, 052507 (2013).

[129] J. Janak, Phys. Rev. B **18**, 7165 (1978).

[130] L. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

[131] P. Mori-Sánchez, A. J. Cohen, and W. Yang, J. Chem. Phys. **125**, 201102 (2006).

[132] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, J. Chem. Phys. **126**, 104102 (2007).

[133] O. A. Vydrov, G. E. Scuseria, and J. P. Perdew, J. Chem. Phys. **126**, 154109 (2007).

[134] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Science **321**, 792 (2008).

[135] R. Haunschild, T. M. Henderson, C. A. Jiménez-Hoyos, and G. E. Scuseria, J. Chem. Phys. **133**, 134116 (2010).

[136] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Chem. Rev. **112**, 289 (2012).

[137] M. Srebro and J. Autschbach, J. Phys. Chem. Lett. **3**, 576 (2012).

[138] J. D. Gledhill, M. J. G. Peach, and D. J. Tozer, J. Chem. Theory Comp. **9**, 4414 (2013).

[139] O. T. Hofmann, V. Atalla, N. Moll, P. Rinke, and M. Scheffler (2013), arXiv:1310.2097v1.

[140] R. Parr and L. Bartolotti, J. Phys. Chem. **87**, 2810 (1983).

[141] J. R. Chelikowsky and M. L. Cohen, in *Handbook on Semiconductors*, edited by T. Moss and P. T. Landsberg (Elsevier, Amsterdam, 1992), vol. 1, chap. 3.

[142] G. Onida, L. Reining, and A. Rubio, Reviews of Modern Physics **74**, 601 (2002).

[143] S. Botti, A. Schindlmayr, R. D. Sole, and L. Reining, Rep. Prog. Phys. **70**, 357 (2007).

[144] J. R. Chelikowsky, N. Troullier, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994).

[145] J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad, Phys. Rev. B **50**, 11355 (1994).

[146] L. Kronik, A. Makmal, M. L. Tiago, M. M. G. Alemany, M. Jain, X. Huang, Y. Saad, and J. R. Chelikowsky, Phys. Status Solidi (B) **243**, 1063 (2006).

[147] http://parsec.ices.utexas.edu.

[148] M. Alemany, M. Jain, L. Kronik, and J. Chelikowsky, Phys. Rev. B **69**, 075101 (2004).

[149] A. Natan, A. Benjamini, D. Naveh, L. Kronik, M. Tiago, S. Beckman, and J. Chelikowsky, Phys. Rev. B **78**, 075109 (2008).

[150] D. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[151] R. Godby and I. White, Phys. Rev. Lett. **80**, 3161 (1998).

[152] A. F. Izmaylov and G. E. Scuseria, J. Chem. Phys. **129**, 034101 (2008).

[153] P. Mori-Sánchez, A. J. Cohen, and W. Yang, Phys. Rev. Lett. **100**, 146401 (2008).

[154] O. Madelung, *Semiconductors: data handbook* (Springer, 2004), 3rd ed.

[155] Here and below it is assumed that the ground states of the system of interest, of its anion and of its cation are not degenerate, or that the degeneracy can be lifted by applying an infinitesimal external field

[156] The fact that Eq. (2) includes contributions only from the N_0- and the $N_0 + 1$-states relies on the conjecture that the series $E(N_0)$ for $N_0 \in \mathbb{N}$ is a convex, monotonously decreasing series. In other words, all ionization energies $I(N_0) := E(N_0 - 1) - E(N_0)$ are positive, and higher ionizations are always larger than the lower ones: $I(N_0 - 1) > I(N_0)$. This conjecture, although strongly supported by experimental data, remains without proof, to the best of our knowledge [4, 124, 136].

[157] The calculations were performed for GaAs in the zincblende crystal structure with the experimental lattice constant of $a = 10.684$ Bohr [154]. A numerical precision of 0.02 eV in the reported energy gaps was obtained with a real-space grid spacing of $h = 0.25$ Bohr and an 11x11x11 k-point sampling scheme. The norm-conserving Troullier-Martins pseudopotentials [N. Troullier and J.L. Martins, Phys. Rev. B **43**, 1993 (1991)] for Ga and As were obtained using the APE program [M.J. Oliveira and F. Nogueira, Comput. Phys. Commun. **178**, 524 (2008), M.J. Oliveira, F. Nogueira, and T. Cerqueira, APE – Atomic Pseudopotential Engine, see http://www.tddft.org/programs/APE/], within the scalar-relativistic approximation, with the electronic configurations of $[Ar]4s^2 4p^1 4d^0$ and $[Ar]4s^2 4p^3 4d^0$ for Ga and As, respectively, with $s/p/d$ cutoff radii of 1.8/2.2/2.8 Bohr and 1.8/2.1/2.5 Bohr, using a non-linear core correction [S.G. Louie, S. Froyen, and M. Cohen, Phys. Rev. B **26**, 1738 (1982)] and choosing the d orbital as the local component in the Kleinman-Bylander projection scheme [L. Kleinman and D.M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982)].

[158] We draw special attention to the assumption underlying the derivation leading to Eq. (19): the states with N_0, $N_0 - 1$, and $N_0 + 1$ electrons have to be pure, i.e. non-degenerate, states. Note that some degeneracies can be trivially removed: Degeneracy between energy levels in the two spin channels is removed by introducing an infinitesimal magnetic field; Degeneracy between symmetric k-points in a periodic crystal is removed by an infinitesimal spatial distortion of the crystal structure. However, non-trivial degeneracies, e.g. a simultaneous fractional occupation of d- and s-bands, requires a non-infinitesimal perturbation for their removal. For these cases, the ensemble approach should be generalized to include more than two components in Eqs. (2) and (8), which may affect the expression obtained for Δ. Therefore, for cases of, e.g., metals and semi-metals a more general treatment is needed.